100XPONENTIAL- FEB 2023

Statistical Savvy: Building Data-Driven Minds in Finance

Enterprise Data Training

15+

STATISTICAL CONCEPTS FOR BUSINESS

PREPARED BY: ISAIAH OLATERU

ISAIAH.OLATERU@100XPONENTIAL.COM

Executive Summary

☑ Business ☐ Policy ☑ Finance

Business Problem: To transition into a data-driven organization, we need to develop a “people” competency plan. This project aims to generate synthetic operational data relevant to the finance industry, specifically finance operations and fund administration. The data-set should reflect the day-to-day experiences of trainees (employees of "Name Withheld Organization - N.W.O") as a core component of the customized experiential data competency program. Additionally, the data-set must be designed to illustrate over 15 foundational statistical concepts essential for business decision-making, such as mean, median, mode, normal and bimodal distributions, outliers, and data types.

Project Goal(s): Create data set via traditional data collection methods - operational tasks timing and synthetic data to inform N.W.O core train data set for the Data Foundations enterprise training program.

Key Results:

A high level summary of output data set used to design the data course in summarized in the output table below. Sub attributes here can also be referred to as features .

DATA OUTPUT TABLE

4	25	15+
CORE DATA ATTRIBUTES	SUB ATTRIBUTES	DATA CONCEPTS

This model correctly predicts 7.8 out of 10 patience that will experience a recurrence. This is based on the recall metric which is the most important metric in this use case.

Approach & Data Summary

- First, a real service area in need of improvement was identified- > Accounts Payable Services. The operational and client problem statement was then articulated and agreed upon: it takes too long to process and deposit client payments. This issue was substantiated by complaints collected over a two-week period.

- As a global organization, the payment process was executed across over 20 offices, depending on client jurisdiction and currency. Key questions to be answered included: how long does it take (in minutes/days) to process and deposit a payment? Which locations are the best performers? These business questions were then translated into a data collection table, with core attributes defined as Location, Risks, Payment Priority, and Time-to-Process (cycle time).

- Sub-attributes were also defined. If we take “Risk” as an example, this was further broken down into high, medium, low and based on client risk ratings that impact payment process steps and checks required. Note the publish data set, the risks types were further encoded into Enhanced (high) and standard (medium and low) risks

- Manual data collection was conducted for 35 events across 6 locations—2 each in Europe, North America, and Asia. From the real data collected based on operations, additional data was generated using imputation techniques.

Description	Core Attributes	Example of Subattributes
Payment processing offices	Location	Dublin, Amsterdam, Jersey, Cayman etc.
Client Risk Rating	Risk	Enhanced (ENH) or Standard (STND)
Payment Types based on preagreement	Payment Priority	Urgent , Non Urgent
Processing Time	Time-to-Process (Cycle Time)	10, 20 , 30 minutes

Deliverables

1. The key (publishable), deliverable for the project is the raw data set. Payment Process Optimisation.xls

2. The data contains both real and synthetic data that simulates payment process across 4 key process steps in the finance/find administration space for N.W.O (a real but undisclosed organization)

3. Information like client, processor names have either been masked or removed all together

Path to my GitHub repository to view code related to this summary and get the raw data-set **https://github.com/YNWA-Algo/Payments-Synthetic-Data**